

NAYAK EXHIBIT B

Research

Introducing LLaMA: A foundational, 65-billion-parameter large language model

February 28, 2023

→ Share on Facebook

→ Share on Twitter

Research Areas

Research

UPDATE: We just launched Llama 2 - for more information on the latest see our blog post on Llama 2.

As part of Meta's commitment to open science, today we are publicly releasing LLaMA (Large Language Model Meta AI), a state-of-the-art foundational large language model designed to help researchers advance their work in this subfield of AI. Smaller, more performant models such as LLaMA enable others in the research community who don't have access to large amounts of infrastructure to study these models, further democratizing access in this important, fast-changing field.

Training smaller foundation models like LLaMA is desirable in the large language model space because it requires far less computing power and resources to test new approaches, validate others' work, and explore new use cases. Foundation models train on a large set of unlabeled data, which makes them ideal for fine-tuning for a variety of tasks. We are making LLaMA available at several sizes (7B, 13B, 33B, and 65B parameters) and also sharing a LLaMA model card that details how we built the model in keeping with our approach to Responsible AI practices.

Over the last year, large language models — natural language processing (NLP) systems with billions of parameters — have shown new capabilities to generate creative text, solve mathematical theorems, predict protein structures, answer reading comprehension questions, and more. They are one of the clearest cases of the substantial potential benefits AI can offer at scale to billions of people.

Even with all the recent advancements in large language models, full research access to them remains limited because of the resources that are required to train and run such large models. This restricted access has limited researchers' ability to understand how and why these large language models work, hindering progress on efforts to improve their robustness and mitigate known issues, such as bias, toxicity, and the potential for generating misinformation.

Smaller models trained on more tokens — which are pieces of words — are easier to retrain and fine-tune for specific potential product use cases. We trained LLaMA 65B and LLaMA 33B on 1.4 trillion tokens. Our smallest model, LLaMA 7B, is trained on one trillion tokens.

Like other large language models, LLaMA works by taking a sequence of words as an input and predicts a next word to recursively generate text. To train our model, we chose text from the 20 languages with the most speakers, focusing on those with Latin and Cyrillic alphabets.

There is still more research that needs to be done to address the risks of bias, toxic comments, and hallucinations in large language models. Like other models, LLaMA shares these challenges. As a foundation model, LLaMA is designed to be versatile and can be applied to many different use cases, versus a fine-tuned model that is designed for a specific task. By sharing the code for LLaMA, other researchers can more easily test new approaches to limiting or eliminating these problems in large language models. We also provide in the paper a set of evaluations on benchmarks evaluating model biases and toxicity to show the model's limitations and to support further research in this crucial area.

To maintain integrity and prevent misuse, we are releasing our model under a noncommercial license focused on research use cases. Access to the model will be granted on a case-by-case basis to academic researchers; those affiliated with organizations in government, civil society, and academia; and industry research laboratories around the world. People interested in applying for access can find the link to the application in our research paper.

We believe that the entire AI community — academic researchers, civil society, policymakers, and industry — must work together to develop clear guidelines around responsible AI in general and responsible large language models in particular. We look forward to seeing what the community can learn — and eventually build — using LLaMA.

[Read the paper](#)[Read the model card](#)[Apply for access to LLaMA](#)

Our approach

[About AI at Meta](#)
[People](#)
[Careers](#)

Product experiences

[Meta AI](#)
[AI Studio](#)

Foundational models

Research

[Infrastructure](#)
[Resources](#)
[Demos](#)

Latest news

[Blog](#)
[Newsletter](#)

Q Search AI content



CONFIDENTIAL

Llama

[Privacy Policy](#) [Terms](#) [Cookies](#)

Meta © 2025